W64–W71 *Nucleic Acids Research, 2014, Vol. 42, Web Server issue* *Published online 29 April 2014*
doi: 10.1093/nar/gku318

NEP: web server for epitope prediction based on antibody neutralization of viral strains with diverse sequences

Gwo-Yu Chuang[1], David Liou[2], Peter D. Kwong[1] and Ivelin S. Georgiev[1,*]

[1]Vaccine Research Center and [2]Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Received January 31, 2014; Revised March 25, 2014; Accepted April 2, 2014

ABSTRACT

Delineation of the antigenic site, or epitope, recognized by an antibody can provide clues about functional vulnerabilities and resistance mechanisms, and can therefore guide antibody optimization and epitope-based vaccine design. Previously, we developed an algorithm for antibody-epitope prediction based on antibody neutralization of viral strains with diverse sequences and validated the algorithm on a set of broadly neutralizing HIV-1 antibodies. Here we describe the implementation of this algorithm, NEP (Neutralization-based Epitope Prediction), as a web-based server. The users must supply as input: (i) an alignment of antigen sequences of diverse viral strains; (ii) neutralization data for the antibody of interest against the same set of antigen sequences; and (iii) (optional) a structure of the unbound antigen, for enhanced prediction accuracy. The prediction results can be downloaded or viewed interactively on the antigen structure (if supplied) from the web browser using a JSmol applet. Since neutralization experiments are typically performed as one of the first steps in the characterization of an antibody to determine its breadth and potency, the NEP server can be used to predict antibody-epitope information at no additional experimental costs. NEP can be accessed on the internet at http://exon.niaid.nih.gov/nep.

INTRODUCTION

The determination of epitopes targeted by antibodies is useful for understanding virus escape (1), antibody optimization (2,3) and epitope-based design of vaccines (4). Structure determination (by, e.g. X-ray crystallography) of antibody–antigen complexes can provide epitope information at the atomic level (5), but in many instances, atomic-level complex structures can be challenging to obtain. Additional experimental methods for epitope delineation are also available, although they are characterized with lower accuracy and typically require substantial experimental effort (5–7). Computational methods for epitope prediction have traditionally aimed at predicting antigen residues that could be part of any antibody epitope, and are thus not antibody specific (8–11). More recently, computational methods for antibody-specific epitope prediction (the prediction of the epitope targeted by an antibody of interest) have been developed (7,12–15). Specifically, we and others have focused on combining antibody–antigen neutralization data with antigen sequence information in order to predict residues that may be part of the epitope for antibodies of interest (7,12,13).

Antibody neutralization assays, which measure the reduction of viral infectivity mediated by antibody, are often performed as one of the first steps in the characterization of an antibody to determine its breadth and potency. Previously, we developed a neutralization-based epitope prediction method that is applicable to antigens that exhibit substantial sequence diversity, such as human immunodeficiency virus 1 (HIV-1) and influenza (7). The algorithm, named NEP for neutralization-based epitope prediction, is based on the premise that sequence variation of epitope residues is more likely to have an effect on antibody neutralization than variation of non-epitope residues. For each antigen residue position, NEP estimates the association between sequence variation and changes in antibody neutralization for a given set of diverse viral strains. A structure of the unbound antigen, if available, can be used for further improvement in the prediction accuracy. NEP has been validated on a set of HIV-1 antibodies targeting a number of different epitopes on the virus: both for retrospective epitope prediction [for 19 antibodies with known complex structures, with a true positive (TP) rate of 0.403 at a 0.05 false positive (FP) rate level] and for prospective epitope prediction (for HIV-1 antibody 8ANC195, with a previously uncharacterized epitope) (7). Similar methods for

*To whom correspondence should be addressed. Tel: +1 301 435 8632; Fax: +1 301 480 2658; Email: ivelin.georgiev@nih.gov

neutralization-based antibody-epitope prediction were also described recently (12,13).

In this paper, we describe the implementation of the NEP algorithm as a web-based server. The NEP server allows the user to predict the epitope for an antibody by using antigen sequence alignment for diverse viral strains, antibody–antigen neutralization data over the same set of strains and (optionally) a structure of the unbound antigen. The results can be downloaded or viewed interactively in a web browser via the JSmol Applet. NEP is the first publicly available server for antibody-epitope prediction using antigen structure and neutralization data of diverse viral strains.

MATERIALS AND METHODS

Epitope-prediction algorithm

For each residue position in an antigen, the NEP algorithm computes a mutual information score (16) between amino acid variation at that position and changes in sensitivity to virus neutralization. Two method variants were implemented in this server, based on our previously published study (7).

(i) 'Neutralization + sequence': each antigen residue i is ranked by the normalized mutual information between amino acid types and neutralization IC_{50} values. The score for residue i is defined as follows:

$$\text{Score}_i = n\text{MI}(X_i; Y) = \frac{\text{MI}(X_i; Y)}{H(X_i)} = \frac{\sum_{y\in Y}\sum_{x\in X_i} p(x, y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)}{-\sum_{x\in X_i} p(x)\log(p(x))},$$

where X_i is a variable that covers the possible amino acid types at position i (the 20 natural amino acid types and a gap in the sequence alignment). Y is a binary variable defined by a user-specified IC_{50} cutoff value that divides strains into a resistant and a sensitive class. $\text{MI}(X_i; Y)$ is the standard mutual information (16) defined between X_i and Y, and $H(X_i)$ is the Shannon entropy of the amino acid types at each residue position (17).

(ii) 'Neutralization + sequence + structure' (if an unbound antigen structure is supplied by the user): the solvent accessible area of each residue is calculated using NACCESS (18); residues with solvent accessible area of less than 5 $Å^2$ are considered as non-solvent-accessible and are excluded from further analysis. The score for each of the remaining residues is then taken to be the normalized mutual information of the residue as calculated in (1) plus a distance-dependent sum of the normalized mutual information of surrounding (in space) residues:

$$\text{Score}_i = n\text{MI}(X_i; Y) + 0.031 * \text{PS}_i + \sum_{j\in A; j\neq i; r_{ij}<9}\left(\left(1-\frac{r_{ij}}{9}\right)*0.8+0.2\right)\left(n\text{MI}(X_j; Y) + 0.031 * \text{PS}_j\right),$$

where A is the set of all antigen solvent-accessible residues and r_{ij} is the shortest distance (in Å) between any pair of side-chain heavy atoms (or Cα for glycine) for residues i and j. An epitope propensity score (PS), a score that estimates the likelihood for each amino acid type to be part of an antibody epitope (11), is also incorporated in the scoring function as a weighted term.

For each calculated residue score, Score_i, an empirical P-value (P_i) is determined by performing permutation tests similar to those performed in (19) and (20). Briefly, the score for each residue position is recalculated 500 times using the same scoring function while permuting the residue types across the viral strains, resulting in 500 scores: $S_{i,p}$, $p \in \{1,2,\ldots,500\}$. The empirical P-value can thus be calculated as $P_i = \sum_{p=1}^{500} I(S_{i,p} \geq \text{Score}_i)/500$, where $I()$ is the indicator function. For example, if $Score_i$ is greater than 495 $S_{i,p}$ but less than 5 $S_{i,p}$ from the permutation test, then P_i is $5/500 = 0.01$.

Server implementation

The epitope prediction protocol consists of a combination of Java code, Perl code, and Linux shell scripts. MAFFT version 7 (21) (FFT-NS-2 method) was incorporated for performing multiple sequence alignment. The server was developed using the Google Web Toolkit (GWT) v2.5 framework, hosted by an Apache Tomcat 6.0 application server running on Red Hat Linux 4. Additionally, the client utilizes the JSmol javascript/HTML5 libraries for molecular structure visualizations. The application architecture is primarily a two-tier system consisting of the client front end and a logical tier comprised of shell-based processes for analysis. Supported platforms for the NEP Server are Chrome, Safari, Firefox and Internet Explorer 9+. More information on GWT may be found at http://www.gwtproject.org/ and on JSmol at http://jsmol.sourceforge.net/.

Comparison with other servers

The prediction accuracy of the two method variants implemented as part of the NEP server was compared to two web servers for identifying specificity-determining positions, SPEER (19) and SigniSite (22). A set of 19 HIV-1 antibodies with available antibody–antigen structures and neutralization data on a panel of ∼180 viral strains (7) was used for the comparison. The true epitope residues were defined as antigen residues for which any heavy atom was within 5 Å of any heavy atom in the given antibody, as observed in the structures. The gp160 amino acid sequences for the different strains were aligned using HIVAlign (http://www.hiv.lanl.gov/content/sequence/VIRALIGN/viralign.html). Since one of the two NEP method variants requires a structure of the unbound antigen as input, only antigen residues present in the complex structure for the respective antibody were used in the analysis (7). For SPEER, the relative entropy term for each residue position was used to rank the predictions. For both NEP and SPEER, a threshold of 50 μg/ml was used to classify the neutralization potency into two classes. For SigniSite, which assigns a Z-score for each residue type at each

position, the maximum of the absolute Z-scores was chosen as the score for each residue position. A two-tailed paired *t*-test was used to evaluate the significance of the differences in prediction accuracy (based on the TP rates at the 0.05 FP rate level) between the different methods; computed *P*-values were further adjusted using the Bonferroni method implemented in R (http://www.r-project.org/) (a total of six *t*-tests were performed). A *P*-value threshold of 0.01 was used to determine if the prediction accuracy of one method was ($P < 0.01$) or was not ($P >= 0.01$) significantly different from another method.

RESULTS

Using the server

The workflow for using the NEP server is shown in Figure 1. The required input is: 1) antibody–antigen neutralization data on a set of diverse viral strains and 2) the sequence alignment for the same set of strains. The users have the option to upload the files directly or copy and paste the content to an input text box. For neutralization data input, the antibody neutralization potency (e.g. IC_{50}) values against different antigen strains must be supplied. Our previous analyses on HIV-1 antibody-epitope prediction demonstrate that strain diversity plays an important role in prediction accuracy (7), so the user should try to increase strain diversity in the input neutralization panel. The users must also specify a threshold value that classifies all the input neutralization values into two classes (sensitive and resistant). For example, if an IC_{50} of 50 μg/ml is specified as the threshold, then all neutralization values of less than 50 μg/ml will be considered sensitive, while all neutralization values equal to or greater than 50 μg/ml will be considered resistant. The antigen sequence alignment of all the strains in the neutralization input must also be supplied, in FASTA format (http://www.ncbi.nlm.nih.gov/BLAST/blastcgihelp.shtml). Since the sequence length of the different strains will likely vary, a reference strain must also be selected for specifying residue numbering in the output. If the input sequences are not aligned, the user can request the server to generate a multiple sequence alignment using MAFTT version 7 (FFT-NS-2 method) (21). Once the input is supplied, the users select one of the two method variants for prediction. If the 'neutralization + sequence + structure' method is selected, a structure of the unbound antigen [in PDB format (23)] must be provided. The residue numbering in the PDB file has to correspond to the residue numbering of the selected reference strain, and the analysis will only be performed for the residues found in the antigen structure file.

For typical problems, NEP requires generally ∼1 (without supplying antigen structure) or 2 min (when supplying antigen structure) for the analysis. Based on the user's input and method selection, NEP ranks each input antigen residue using the method's scoring function. Higher scores indicate higher likelihood that the respective residue is part of the epitope for the given antibody. An empirical *P*-value based on a permutation test (see the Materials and Methods section) is also provided for each residue score. Two different residue numbering systems are displayed in the output: (i) the residue numbering in the alignment (gaps included) and (ii) the residue numbering of the reference strain (gaps excluded). The user also has the option to download the results in a CSV (Comma-Separated Values) format and the multiple sequence alignment file in FASTA format. The residue median score and 95-percentile score are also computed. Our previous analysis suggested that there is a significant relationship between residue score distribution and prediction accuracy for a given epitope: the larger the ratio of the 95-percentile score to the median score, the higher the prediction accuracy (7). The server therefore implements a warning of low prediction confidence when this ratio is less than a threshold (five for the 'neutralization + sequence' method and two for the 'neutralization + sequence + structure' method). When using the 'neutralization + sequence + structure' method, the antigen structure is displayed in the JSmol viewer. The user can map top-scoring residues on the antigen structure by checking on or off residues from the control panel. The user can further adjust the structural appearance using the JSmol functionalities.

Case studies

Epitope prediction for HIV-1 antibody VRC01. VRC01 is a broadly neutralizing HIV-1 antibody targeting the receptor binding site on the HIV-1 Env glycoprotein (24–26). Based on VRC01 neutralization data and sequence alignment of 181 HIV-1 strains (7), and using an unbound gp120-core antigen structure as input [PDB:3TGT (27)], all top nine residues predicted by NEP (residues 279–280, 456–461, 466; Figure 2) were part of the VRC01 epitope as defined by the gp120/VRC01 complex structure [PDB:3NGB, (28)] (with at least one antigen heavy atom within 5 Å of any antibody heavy atom).

Epitope prediction for HIV-1 antibody PG9. Antibody PG9 is a member of a group of glycopeptide-reactive antibodies that targets the V1/V2 region of the HIV-1 Env glycoprotein and potently neutralizes the majority of circulating strains of HIV-1 (29). Based on PG9 neutralization data and sequence alignment of 174 HIV-1 strains (7), and using the V1/V2 region of a scaffolded V1/V2 structure [PDB:3U4E (30)] as antigen structure input, all top six residues predicted by NEP (residues 160, 162–163, 169–171) were part of the PG9 epitope as defined by the scaffolded V1/V2–PG9 complex structure [PDB:3U4E (30)]. Specifically, NEP ranked residue 160 as the top predicted epitope residue (Figure 3), in agreement with the known dependence of PG9 neutralization on the presence of N-linked glycan at residue position 160 (30,31).

More case studies can be found online at http://exon.niaid.nih.gov/nep/#examples.

Comparison with other servers

We compared the epitope prediction accuracy of our methods to two state-of-the-art web servers, SPEER (19) and SigniSite (22), that are not specifically designed for predicting antibody epitopes but provide the ability to identify specificity-determining positions within a protein. For the comparison, we used a benchmark set of 19 HIV-1 antibodies used in our previous study (7). In particular, we focused on the TP rate at low FP rate levels (specifically, at an

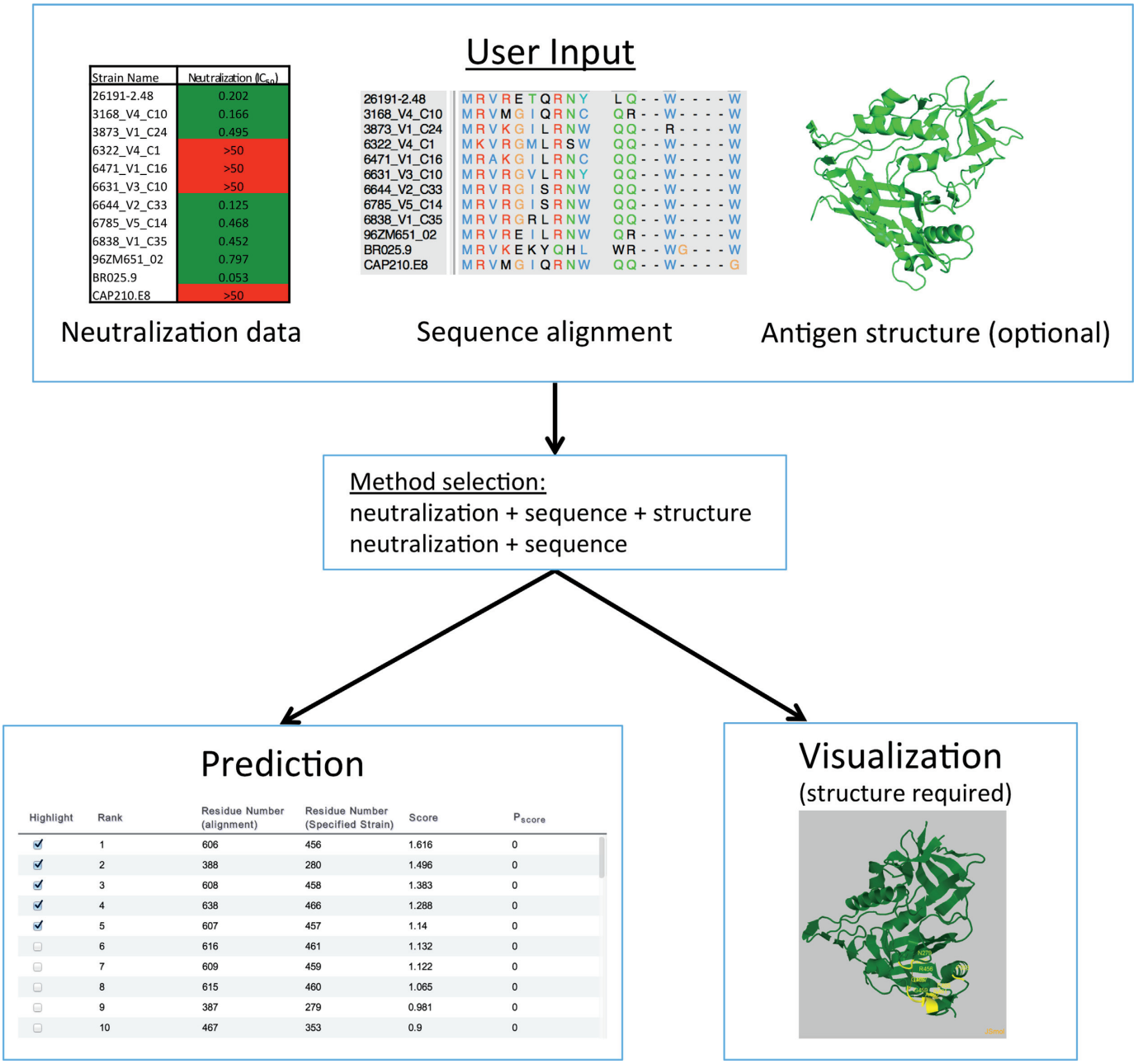


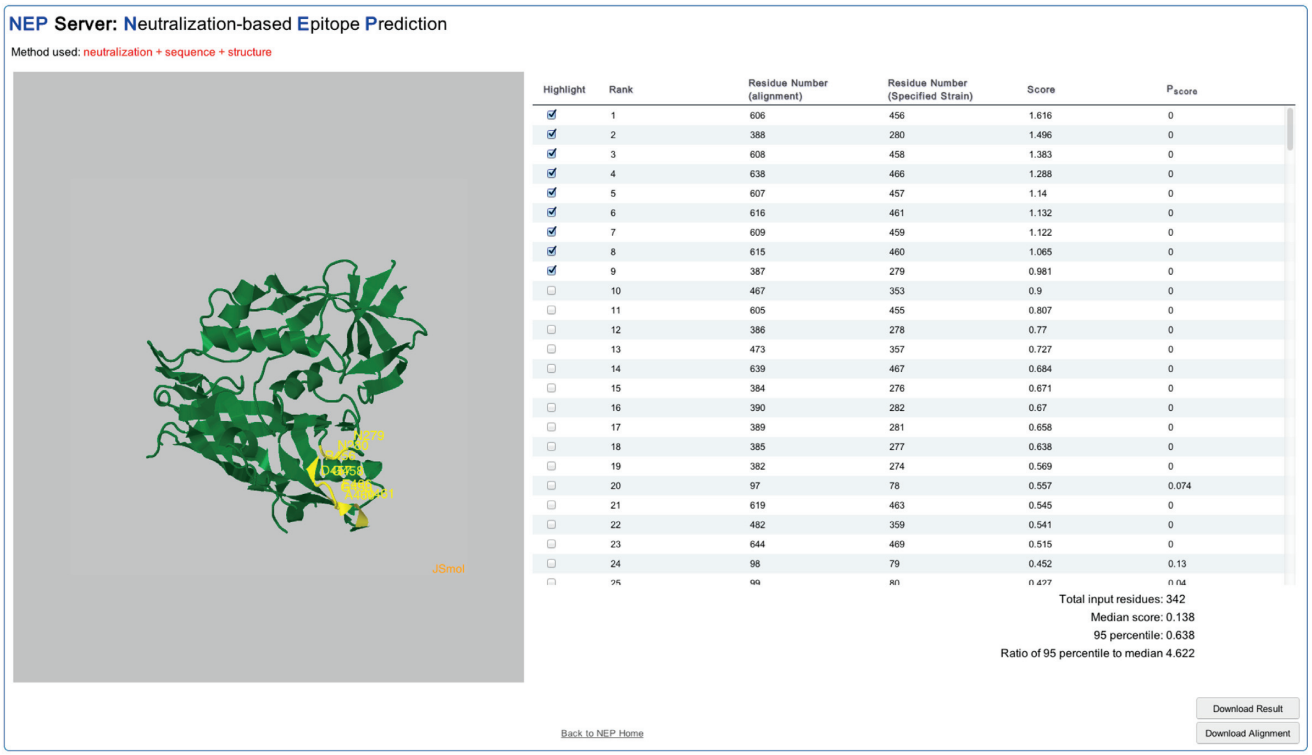
Figure 1. Overview of the NEP server workflow. The user first inputs the antibody neutralization data, the sequence alignment of the antigen strains and (optionally) an unbound structure of the antigen. Next, one of the two method variants for epitope prediction is selected. The result of the analysis is output as a table with residues ranked according to their NEP score, and the prediction can be visualized on the antigen structure (if provided).

FP rate = 0.05). The 'neutralization + sequence' NEP variant did not have a significantly different accuracy compared to the other two servers (Figure 4). The 'neutralization + sequence + structure' NEP variant, however, significantly outperformed both other servers ($P < 0.01$).

DISCUSSION

The NEP server predicts antibody epitopes based on antibody neutralization data of diverse viral strains. Without supplying the antigen structure (the 'neutralization + sequence' method variant), the algorithm aims at identifying residue positions where amino acid variation is associated with changes in neutralization potency. In principle, this method predicts the 'functional epitope' for an antibody rather than the 'structural epitope', and can be used to identify resistance mutations. One limitation of this approach is that for residue positions that are conserved in the sequence alignment, the mutual information between amino acid variation at the given residue position and the neutralization data is zero. As a result, the method is not able to identify conserved residue positions as epitope residues. In contrast, when the antigen structure is supplied (the 'neutralization + sequence + structure' method

(a)



(b)

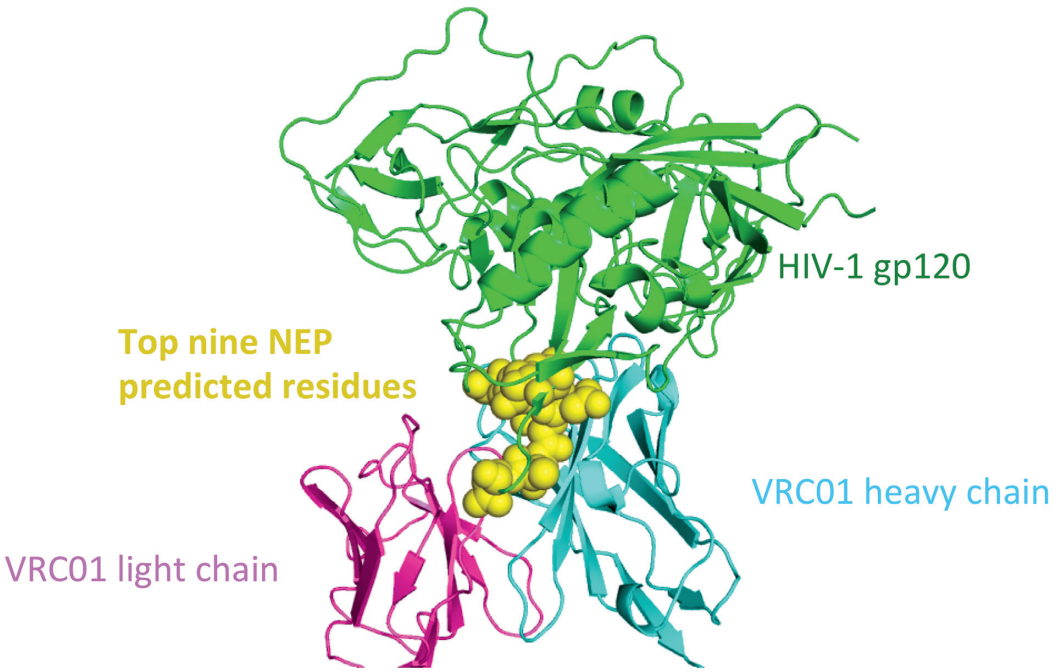
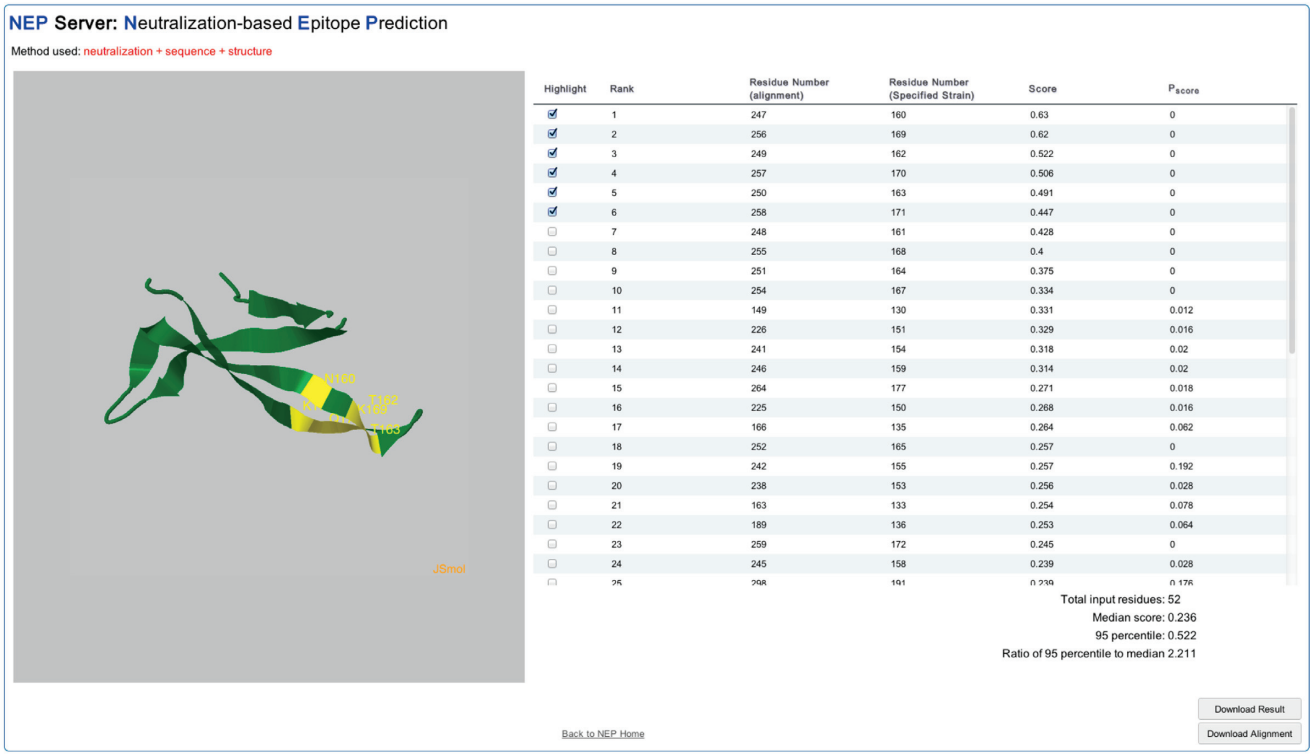


Figure 2. Epitope prediction for HIV-1 antibody VRC01. (**a**) Predictions from the NEP server, with the top nine residues highlighted in yellow on a structure of the antigen (an unliganded HIV-1 gp120 core, PDB:3TGT). (**b**) Crystal structure of a gp120/VRC01 complex (PDB:3NGB), with the HIV-1 gp120 antigen in green, the VRC01 heavy chain in cyan and the VRC01 light chain in magenta. The top nine residues predicted by NEP (yellow spheres) are all part of the VRC01 epitope as defined by the crystal structure.

(a)



(b)

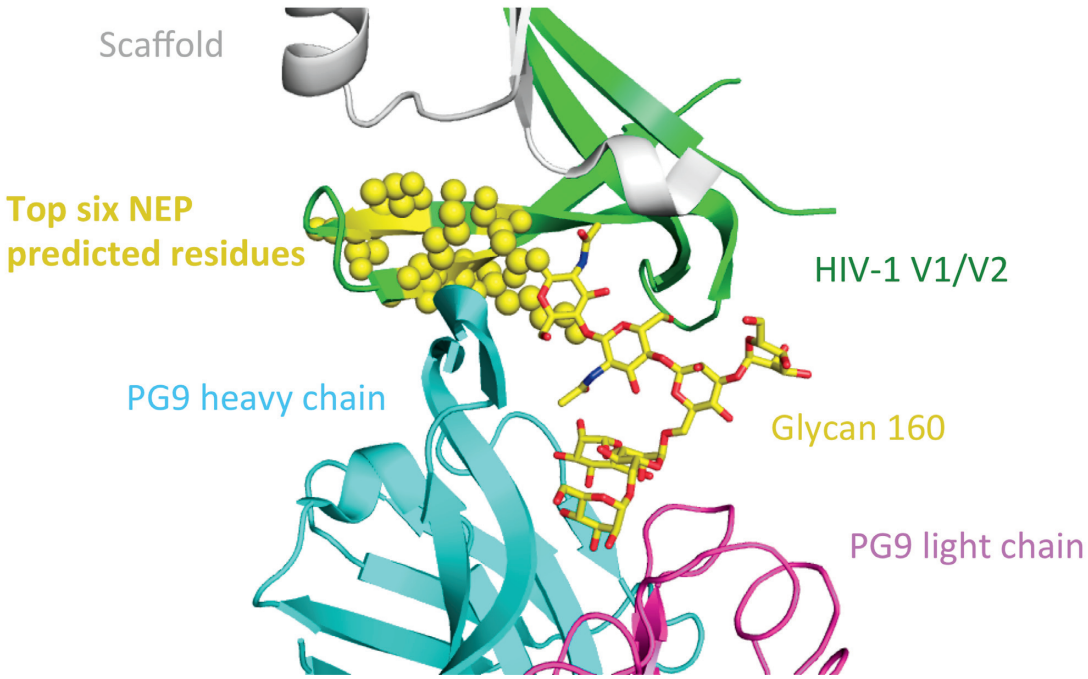


Figure 3. Epitope prediction for HIV-1 antibody PG9. (**a**) Predictions from the NEP server, with the two top-scoring residues 160 and 169 highlighted in yellow on a structure of the antigen [residue 126–196 of a scaffolded HIV-1 V1/V2 (PDB:3U4E)]. (**b**) Crystal structure of a scaffolded HIV-1 gp120 V1/V2 region in complex with PG9 (PDB:3U4E), with the HIV-1 gp120 V1/V2 region in green, the PG9 heavy chain in cyan and the PG9 light chain in magenta. The top six residues predicted by NEP (yellow spheres) are all part of the PG9 epitope as defined by the crystal structure. Glycan-160 (stick representation) makes contact with both the heavy and light chains of PG9.

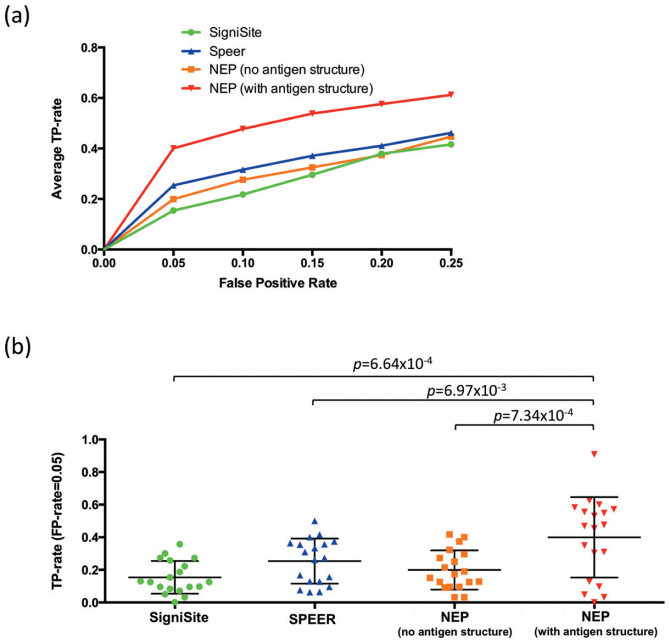


Figure 4. Comparison of epitope prediction accuracy for NEP versus other methods. The accuracy of the NEP server was compared to SigniSite and Speer, based on a benchmark set of 19 HIV-1 antibodies. (**a**) The average TP-rates at different FP-rate levels (0.05, 0.10, 0.15, 0.20 and 0.25) for each method. (**b**) The TP-rate at an FP-rate level of 0.05 for each of the 19 antibodies, with the average (long black horizontal bar) and standard error (black vertical bar) shown for each method; *P*-values of less than 0.01 are displayed on top.

variant), the algorithm aims to predict the antibody structural epitope based on the assumption that epitope residues are more likely to cluster, rather than be isolated, on the antigen surface. Thus, with 'neutralization + sequence + structure', a conserved residue could still be predicted as an epitope residue if there are many other high-scoring spatially proximal residues. The improved prediction accuracy from adding the antigen structure as input was also evident when comparing NEP with SPEER and SigniSite: without supplying a structure of the antigen, the performance of NEP was not significantly different from the performance of SPEER or SigniSite, which use different techniques to quantify the relationship between residue type at each position and neutralization potency; when a structure of the antigen was supplied as input, NEP significantly outperformed the other two servers. To address the potential issue that the computation of the mutual information scores may be affected by the observed amino acid diversity at each residue position, we (i) normalized the mutual information score with the Shannon entropy of the amino acid composition at each residue position and (ii) calculated an empirical *P*-value based on permutation tests to give a statistical confidence for each score.

The NEP server was designed to extract antibody-epitope information from neutralization data. The NEP algorithm in theory, however, could also extract information from other sources of quantitative functional data, such as binding affinity (e.g. K_D values). In addition to antibody-epitope prediction, the server could also potentially be used for the identification of small-molecule binding sites and resistance mutations, provided that such functional data on diverse viral strains are available. NEP should also be applicable to antibody-epitope prediction for other viruses that exhibit substantial sequence diversity, such as influenza, and can be of general utility for efficient initial screening of antibody epitopes.

CONCLUSIONS

We have implemented a web-server, NEP, that utilizes antibody neutralization potency data of diverse viral strains to predict epitope residues. In addition to antigen sequence information, the users have the option of supplying a structure of the free antigen, for improved prediction accuracy. The results from NEP can be downloaded from or viewed interactively in the web-browser. NEP is publicly accessible with no registration or login requirements.

ACKNOWLEDGMENTS

We thank the Structural Biology Section and Structural Bioinformatics Core at the Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health for helpful discussions and comments on the manuscript. We also thank Kui Shen from the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, for suggestions on statistical analyses.

FUNDING

Vaccine Research Center's Intramural Research Program, National Institute of Allergy and Infectious Diseases; Office of AIDS Research, National Institutes of Health (NIH). Funding for open access charge: Vaccine Research Center's Intramural Research Program, National Institute of Allergy and Infectious Diseases; NIH.
Conflict of interest statement. None declared.

REFERENCES

1. Adams,O., Bonzel,L., Kovacevic,A., Mayatepek,E., Hoehn,T. and Vogel,M. (2010) Palivizumab-resistant human respiratory syncytial virus infection in infancy. *Clin. Infect. Dis.*, **51**, 185–188.
2. Diskin,R., Scheid,J.F., Marcovecchio,P.M., West,A.P. Jr, Klein,F., Gao,H., Gnanapragasam,P.N., Abadir,A., Seaman,M.S., Nussenzweig,M.C. *et al.* (2011) Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science*, **334**, 1289–1293.
3. Yu,C.M., Peng,H.P., Chen,I.C., Lee,Y.C., Chen,J.B., Tsai,K.C., Chen,C.T., Chang,J.Y., Yang,E.W., Hsu,P.C. *et al.* (2012) Rationalization and design of the complementarity determining region sequences in an antibody-antigen recognition interface. *PloS ONE*, **7**, e33340.
4. Ofek,G., Guenaga,F.J., Schief,W.R., Skinner,J., Baker,D., Wyatt,R. and Kwong,P.D. (2010) Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 17880–17887.
5. Gershoni,J.M., Roitburd-Berman,A., Siman-Tov,D.D., Freund,N.T. and Weiss,Y. (2007) *Epitope Mapping : The First Step in Developing Epitope-Based Vaccines (Drug Development)*. Adis International, Auckland, New Zealand.
6. Geysen,H.M., Meloen,R.H. and Barteling,S.J. (1984) Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 3998–4002.

7. Chuang,G.Y., Acharya,P., Schmidt,S.D., Yang,Y., Louder,M.K., Zhou,T., Kwon,Y.D., Pancera,M., Bailer,R.T., Doria-Rose,N.A. *et al.* (2013) Residue-level prediction of HIV-1 antibody epitopes based on neutralization of diverse viral strains. *J. Virol.*, **87**, 10047–10058.
8. Parker,J.M., Guo,D. and Hodges,R.S. (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, **25**, 5425–5432.
9. Kulkarni-Kale,U., Bhosle,S. and Kolaskar,A.S. (2005) CEP: a conformational epitope prediction server. *Nucleic Acids Res.*, **33**, W168–W171.
10. Ponomarenko,J., Bui,H.H., Li,W., Fusseder,N., Bourne,P.E., Sette,A. and Peters,B. (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, **9**, 514.
11. Haste Andersen,P., Nielsen,M. and Lund,O. (2006) Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci*, **15**, 2558–2567.
12. West,A.P. Jr, Scharf,L., Horwitz,J., Klein,F., Nussenzweig,M.C. and Bjorkman,P.J. (2013) Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc. Natl. Acad. Sci. U.S.A*, **110**, 10598–10603
13. Ferguson,A.L., Falkowska,E., Walker,L.M., Seaman,M.S., Burton,D.R. and Chakraborty,A.K. (2013) Computational prediction of broadly neutralizing HIV-1 antibody epitopes from neutralization activity data. *PloS ONE*, **8**, e80562.
14. Zhao,L., Wong,L. and Li,J. (2011) Antibody-specified B-cell epitope prediction in line with the principle of context-awareness. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **8**, 1483–1494.
15. Ahmad,S. and Mizuguchi,K. (2011) Partner-aware prediction of interacting residues in protein-protein complexes from sequence data. *PloS ONE*, **6**, e29104.
16. Steuer,R., Kurths,J., Daub,C.O., Weise,J. and Selbig,J. (2002) The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics*, **18 (Suppl.2)**, S231–S240.
17. Shunsuke,I. (1993) *Information Theory For Continuous Systems*. World Scientific Publishing Co., Singapore.
18. Hubbard,S.J.T.J.M. (1993) *'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology*. University College London, London.
19. Chakraborty,A., Mandloi,S., Lanczycki,C.J., Panchenko,A.R. and Chakrabarti,S. (2012) SPEER-SERVER: a web server for prediction of protein specificity determining sites. *Nucleic Acids Res.*, **40**, W242–W248.
20. Shen,K. and Tseng,G.C. (2010) Meta-analysis for pathway enrichment analysis when combining multiple genomic studies. *Bioinformatics*, **26**, 1316–1323.
21. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
22. Jessen,L.E., Hoof,I., Lund,O. and Nielsen,M. (2013) SigniSite: identification of residue-level genotype-phenotype correlations in protein multiple sequence alignments. *Nucleic Acids Res.*, **41**, W286–W291.
23. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
24. Wu,X., Yang,Z.Y., Li,Y., Hogerkorp,C.M., Schief,W.R., Seaman,M.S., Zhou,T., Schmidt,S.D., Wu,L., Xu,L. *et al.* (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*, **329**, 856–861.
25. Zhou,T., Xu,L., Dey,B., Hessell,A.J., Van Ryk,D., Xiang,S.H., Yang,X., Zhang,M.Y., Zwick,M.B., Arthos,J. *et al.* (2007) Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature*, **445**, 732–737.
26. Dey,B., Svehla,K., Xu,L., Wycuff,D., Zhou,T., Voss,G., Phogat,A., Chakrabarti,B.K., Li,Y., Shaw,G. *et al.* (2009) Structure-based stabilization of HIV-1 gp120 enhances humoral immune responses to the induced co-receptor binding site. *PLoS Pathog.*, **5**, e1000445.
27. Kwon,Y.D., Finzi,A., Wu,X., Dogo-Isonagie,C., Lee,L.K., Moore,L.R., Schmidt,S.D., Stuckey,J., Yang,Y., Zhou,T. *et al.* (2012) Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 5663–5668.
28. Zhou,T., Georgiev,I., Wu,X., Yang,Z.Y., Dai,K., Finzi,A., Kwon,Y.D., Scheid,J.F., Shi,W., Xu,L. *et al.* (2010) Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science*, **329**, 811–817.
29. Walker,L.M., Phogat,S.K., Chan-Hui,P.Y., Wagner,D., Phung,P., Goss,J.L., Wrin,T., Simek,M.D., Fling,S., Mitcham,J.L. *et al.* (2009) Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science*, **326**, 285–289.
30. McLellan,J.S., Pancera,M., Carrico,C., Gorman,J., Julien,J.P., Khayat,R., Louder,R., Pejchal,R., Sastry,M., Dai,K. *et al.* (2011) Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*, **480**, 336–343.
31. Doores,K.J. and Burton,D.R. (2010) Variable loop glycan dependency of the broad and potent HIV-1-neutralizing antibodies PG9 and PG16. *J. Virol.*, **84**, 10510–10521.